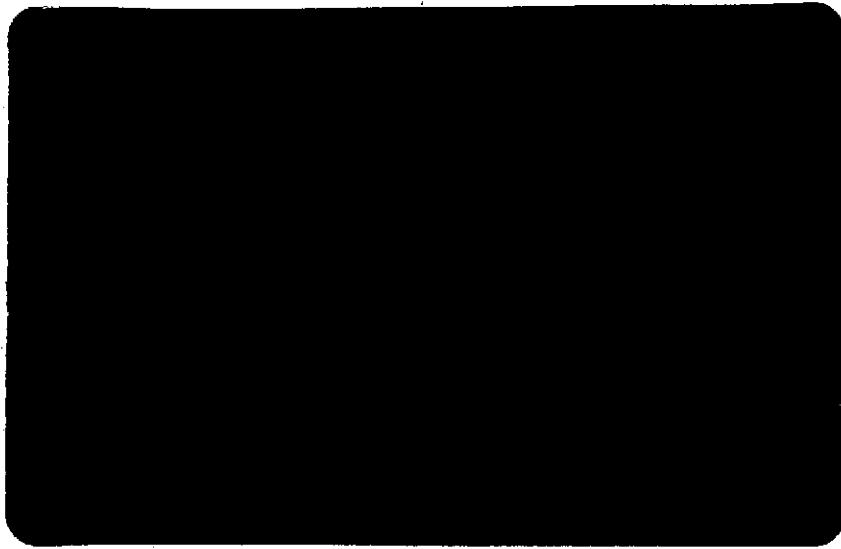


NASA CR-
140297



(NASA-CR-140297) CLASSIFICATION BY MEANS
OF B-SPLINE POTENTIAL FUNCTIONS WITH
APPLICATIONS TO REMOTE SENSING (Rice
Univ.) 9 p HC \$4.00

CSCL 12A

N74-34985

Unclass
51189

G3/19

I C S A

INSTITUTE FOR COMPUTER SERVICES AND APPLICATIONS

RICE UNIVERSITY

Classification by Means of
B-Spline Potential Functions
with Applications to
Remote Sensing
by
J.O. Bennett*, R.J.P. de Figueiredo
and J.R. Thompson
Dept. of Mathematical Sciences
Rice University

ABSTRACT

This paper presents the method of potential functions using B-splines as potential functions in the estimation of likelihood functions (probability density functions conditioned on pattern classes) or of the resulting discriminant functions. Integrated means square consistency of this technique is discussed. Experimental results of using the likelihood functions thus obtained in the classification of remotely sensed data are given.

Institute for Computer Services Applications
Rice University
Houston, Texas

May, 1974

Research supported in part by NASA contract NAS 9-12776
*Now with Esso Production Research, Houston, Texas

CLASSIFICATION BY MEANS OF B-SPLINE POTENTIAL FUNCTIONS WITH APPLICATION TO REMOTE SENSING

John O. Bennett*
Esso Production Research Co.
Houston, Texas

Rui J. P. de Figueiredo†
Rice University
Houston, Texas

James R. Thompson‡
Rice University
Houston, Texas

Abstract

This paper presents the method of potential functions using B-splines as potential functions in the estimation of likelihood functions (probability density functions conditioned on pattern classes) or of the resulting discriminant functions. Integrated mean square consistency of this technique is discussed. Experimental results of using the likelihood functions thus obtained in the classification of remotely sensed data are given.

The method of "potential functions" (also called "kernel functions") for the direct construction of likelihood functions and discriminant functions has been widely discussed in the literature on statistics and pattern classification (see for example [4] and the references therein). In what follows, first we review very briefly this method. Second, we present its integrated-mean-square (IMS) consistency and give a formula for the value of the mesh parameter $h(N)$ (to be defined in section 2) which is optimal with respect to IMS convergence. Next, we discuss the use of multivariate B-splines as potential functions, bringing into the discussion the IMS consistency criteria mentioned above. Finally, we present some of the experimental results obtained when likelihood functions constructed by means of B-spline potential functions were used to classify remotely sensed data pertaining to the Purdue LARS flight line C1.

1. Likelihood Functions and Discriminant Functions in Pattern Classification

As a preamble to our results, let us briefly recall the Bayesian solution to the pattern classification problem.

Suppose that observations made on patterns, which are to be classified as pertaining to one of the pattern classes H^1, \dots, H^M , appear as n -vectors belonging to the real Euclidian space R^n . Then any given observation $x = \text{col}(x_1, \dots, x_n)$ may be viewed as a realization of a random vector $X = \text{col}(X_1, \dots, X_n)$. Associated with each pattern class H^j , $j = 1, \dots, M$, there is the conditional probability density function** $f_X(x/H^j)$, called the likelihood function for the class H^j , and the prior probability H^j for that class. The Bayes decision rule, which minimizes the probability of misclassification, consists of classifying any observed x as arising from H^j if

* Supported by NASA contract NAS-9-12776 while at Rice University.

† Supported by the NSF Grant GK-36375.

‡ Supported by ONR Grant NR-042-283.

** Even though $f(x)$ is the correct notation for the "value" at x of a function f or $f(\cdot)$, we often use the same notation $f(x)$ for the "function" and "function values" when the meaning is clear from the context.

$$P_j f_X(x/H^j) - P_i f_X(x/H^i) > 0, \quad i \neq j, \quad i=1, \dots, n \quad (1)$$

The left side of (1)

$$g_{ji}(x) \equiv P_j f_X(x/H^j) - P_i f_X(x/H^i) \quad (2)$$

is called a discriminant function. Since (1) is equivalent to

$$\tilde{g}_{ji}(x) \equiv \log(f_X(x/H^j)/f_X(x/H^i)) + \log(P_j/P_i) > 0, \quad (1a)$$

$\tilde{g}_{ji}(x)$ is also sometimes called a discriminant function.

For $j = 1, \dots, M$, let there be given the n -vectors $y_1^{(j)} = \text{col}(y_{11}^{(j)}, \dots, y_{1n}^{(j)}), y_2^{(j)}, \dots, y_{N_j}^{(j)} = \text{col}(y_{N_j1}^{(j)}, \dots, y_{N_jn}^{(j)})$ constituting

the training set $T_j(N_j)$ belonging to the pattern class H^j . The problems to which we will be addressing are:

- (a) Given $T_j(N_j)$ construct an estimate $\hat{f}_X(x/H^j, T_j(N_j))$ of $f_X(x/H^j)$;
- (b) Given $T_j(N_j)$ and $T_i(N_i)$ construct an estimate

$$\hat{g}_{ji}(x; T_j(N_j), T_i(N_i)) \text{ of } g_{ji}(x).$$

For simplicity in notation, from now on we will drop the superscript and subscript j whenever it is clear that we are referring to the estimation of a likelihood function pertaining to a given class H^j , and rewrite $f_X(x/H^j)$ and $\hat{f}_X(x/H^j, T_j(N_j))$ simply as $f_X(x)$ and $\hat{f}_X(x/T(N))$, respectively.

2. The Method of Potential Functions

We will now indicate how the method of potential functions is used in the solution of problems (a) and (b) above.

According to this method, in the solution of problem (a), the estimate of $f_X(x/T(N))$ is constructed in the form

$$\hat{f}_X(x/T(N)) = N^{-1} \sum_{k=1}^N \Phi(x, y_k), \quad (3)$$

where $\Phi(x, z)$ called a "potential function" or "kernel function" is a real-valued function of the n -vectors x and z , satisfying appropriate conditions.

For the one-dimensional case, i.e. for x and y_k in \mathbb{R}^1 , Parzen [7] was one of the first investigators to suggest the construction of a probability density function using (3) and for this reason (3) is often called a Parzen estimator of the probability density function $f_X(x)$.

Parzen suggested specifically potential functions of the form

$$\varphi(x, z) = h^{-1}(N)K(h^{-1}(N)(x-z)), \quad (4)$$

where $h(N)$, is a "mesh parameter", dependent on N , sufficiently small so as to validate the assumption of $f_X(x)$ being nearly constant on any interval $(z-h(N), z+h(N))$. Parzen [7] gave conditions on K and $h(N)$, which guarantee mean square consistency of (3) for a wide class of densities. He also gave a formula for optimal $h(N)$, i.e. the values of $h(N)$, $n = 1, 2, \dots$, which maximize the rate of convergence of an approximation to the mean square error to zero.

Parzen's results were extended to the n -dimensional case by Murthy [6] and Cacoullos [1]. Note that in this case, we have in general n scalar mesh parameters $h_1(N), \dots, h_n(N)$ and (4) is replaced by

$$\varphi(x, z) = (h_1(N)h_2(N)\dots h_n(N))^{-1}K(h^{-1}(N)(x-z)), \quad (5)$$

where

$$H(N) = \text{Diag}(h_1(N), \dots, h_n(N)) \quad (6)$$

However, by suitable normalization, we may make all $h_i(N)$, $i = 1, \dots, n$ the same, i.e.

$$h_1(N) = h_2(N) = \dots = h_n(N) = h(N), \quad (7)$$

which when substituted in (5) leads to

$$\varphi(x, z) = h^{-n}(N)K(h^{-1}(N)(x-z)). \quad (8)$$

So from now on, without loss in generality we will assume equation (7) and equation (8) hold.

Referring next to Problem (b), it is clear that the above technique can be used to estimate $g_{ji}(x)$ directly from the training sets $T_j(N_j)$ and $T_i(N_i)$. In fact, let the elements of $T_j(N_j) \cup T_i(N_i)$, ordered in any arbitrary way, be labeled z_k , $k = 1, \dots, N_j + N_i$, and define the function

$$\psi_{ji}(x, z_k) = N_i P_j \varphi(x, z_k) u_j(z_k) - N_j P_i \varphi(x, z_k) u_i(z_k), \quad (9)$$

where, for $\ell = j, i$

$$u_\ell(z_k) = \begin{cases} 1, & \text{if } z_k \in T_\ell(N_\ell) \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Then substituting (3) in (2) and using (9), we obtain

$$\hat{g}_{ji}(x; T_j(N_j), T_i(N_i)) = (N_j N_i)^{-1} \sum_{k=1}^{N_j + N_i} \psi_{ji}(x, z_k) \quad (11)$$

It clearly follows from this definition that our consistency results developed for (3) apply also to (11).

3. Integrated Mean Square Consistency and Optimality

The above-mentioned consistency and optimality results are with respect to the mean square error

$$E\{(\hat{f}_X(x/T(N)) - f_X(x))^2\} \quad (12)$$

for a given x . However, in approximating f_X one is most likely to have some a priori knowledge of the "global behavior", like "global smoothness", of the function to be approximated rather than its "local behavior". In such circumstances, the integrated mean square (IMS) error criterion in the choice of optimal $h(N)$ becomes more meaningful as we shall explain in section 4. The IMS error is defined by

$$V(T(N)) = \int E\{(\hat{f}_X(x/T(N)) - f(x))^2\} dx, \quad (13)$$

where the integration is performed over \mathbb{R}^n .

The following result is proved in the Appendix.

Theorem 1. Suppose

(I) The random samples y_1, \dots, y_N are independently and identically distributed as X whose density is f_X ;

(II) $f_X \in L^2(\mathbb{R}^n)$;

(III) $K: \mathbb{R}^n \rightarrow \mathbb{R}^+$ (where $\mathbb{R}^+ =$ set of nonnegative real numbers) is such that

(III-1) $K \in L^2(\mathbb{R}^n)$,

(III-2) $\int_{\mathbb{R}^n} K(x) dx = 1$,

(III-3) $\text{ess sup}_{x \in \mathbb{R}^n} K(x) < \infty$,

(III-4) $\lim_{\|x\| \rightarrow \infty} \|x\| K(x) = 0$,

where $\|x\| = (x_1^2 + \dots + x_n^2)^{1/2}$, and

(IV) $h(N)$ is such that

(IV-1) $\lim_{N \rightarrow \infty} h(N) = 0$

and

(IV-2) $\lim_{N \rightarrow \infty} N h^2(N) = \infty$.

Then $\hat{f}_X(x/T(N))$ is a consistent estimator of $f_X(x)$ in the IMS error sense, that is

$$\lim_{N \rightarrow \infty} V(T(N)) = 0.$$

We next seek a formula for the value $h^*(N)$ of $h(N)$, which optimizes the IMS rate of convergence, of \hat{f}_X to f_X . This is obtained by modifying Cacoullos' [1] result for the mean square convergence case as follows:

Theorem 2. Let the hypotheses of Theorem 1 hold and assume, in addition, that K is symmetric (i.e. $K(-x) = K(x)$) and f_X is thrice differentiable and such that the second partial derivative of f_X are in $L^2(\mathbb{R}^n)$. Then within $o(h^4(N))$ the IMS error (defined by (13)) is minimized by choosing

$$h^*(N) = N^{-1/(n+4)} \left[\frac{n\|K\|_2^2}{\sum_{i=1}^n \sum_{j=1}^n \mu_{ij} \frac{\partial^2 f_X}{\partial x_i \partial x_j}^2} \right]^{\frac{1}{n+4}}, \quad (14)$$

where, for a function g ,

$$\|g\|_k = \left(\int_{\mathbb{R}^n} |g(x)|^k dx \right)^{1/k} \quad (15)$$

and

$$\mu_{ij} = \int_{\mathbb{R}^n} x_i x_j K(x) dx. \quad (16)$$

The optimal rate of convergence, corresponding to the choice (14), is

$$V^*(T(N)) = (4^{-1} n+1)^{-\frac{n}{n+4}} N^{-\frac{4}{n+4}} \left\| \sum_{i=1}^n \sum_{j=1}^n \mu_{ij} \frac{\partial^2 f_X}{\partial x_i \partial x_j} \right\|_2^2 \left\| \frac{8}{K} \right\|_2^{\frac{n}{n+4}} + o(h^4(N)). \quad (17)$$

If, as we shall assume in the following section, the kernel K has the product form

$$K(x) = \prod_{i=1}^n K_0(x_i), \quad (18)$$

where K_0 is an even one-dimensional kernel, then Theorem 2 further simplifies to:

Theorem 3: Under the condition on K just stated, the results (14) and (17) of the preceding theorem assume the forms (19) and (20) below:

$$h^*(N) = N^{-\frac{1}{n+4}} \left[\frac{n\|K_0\|_2^{2n}}{\sum_{i=1}^n \frac{\partial^2 f_X}{\partial x_i^2}^2} \right]^{\frac{1}{n+4}}, \quad (19)$$

$$V^*(T(N)) = (4^{-1} n+1)^{-\frac{n}{n+4}} N^{-\frac{4}{n+4}} \Lambda_1(f_X) \Lambda_2(K_0), \quad (20)$$

where

$$\sigma_0^2 = \int_{-\infty}^{\infty} x_i^2 K_0(x_i) dx_i, \quad (21)$$

$$\Lambda_1(f_X) = \left\| \sum_{i=1}^n \frac{\partial^2 f_X}{\partial x_i^2} \right\|_2^{\frac{2n}{n+4}} \quad (22)$$

and

$$\Lambda_2(K_0) = \sigma_0^{-\frac{4n}{n+4}} \|K_0\|_2^{\frac{8n}{n+4}} \quad (23)$$

We will call $\Lambda_2(K_0)$ the optimal kernel-dependent rate of convergence since it represents the part of the right side of (20) which depends on K_0 . We omit proofs of theorems 2 and 3 above since they represent straightforward extensions of those in [1].

4. B-Splines as Potential Functions

It is clear from the formulas in (19) and (20) that the choice of the optimal $h^*(N)$, and hence $V^*(T(N))$, depends on the properties of f_X and on the structure of the product kernel K .

The L^2 norm of the second derivative of a function represents a measure of the "global smoothness" of the function. In this sense, according to (22), $\Lambda_1(f_X)$ is a monotonically increasing function of the "global smoothness" of f_X . Thus an a priori knowledge of the smoothness can be incorporated in the formulas (19), (22), and (20) by assigning a value to

$$\left\| \sum_{i=1}^n \frac{\partial^2 f_X}{\partial x_i^2} \right\|_2^2$$

in those formulas.

In picking K one would like to choose its structure so that the optimal kernel-dependent rate of convergence $\Lambda_2(K_0)$ is minimized. Then the choice of the support of K_0 represents a compromise between the minimization of its second moment and its L^2 norm. K_0 must also be at least as smooth as f_X , particularly if only a few training samples are available. Based on these considerations, we suggest for the structure of K_0 a univariate B-spline, and hence for K a product of n such splines. Such a choice for K_0 constitutes a compromise between a Gaussian kernel and a square kernel

$$K_0(x_i) = \begin{cases} \frac{1}{2}, & |x_i| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

We note that a multivariate polynomial, as suggested by Specht [10], while certainly adequate for approximating a large class of discriminant functions given in the form (1a), is certainly unsuitable for the representation of K over the entire \mathbb{R}^n since it violates the conditions (III) of Theorem 1.

Let any given component variable x_i of x be denoted by ξ . Then a univariate B-spline of degree $m-1$ with support on (ξ_0, ξ_m) and knots

$$\xi_0 < \xi_1 < \dots < \xi_m, \quad (25)$$

is defined by [2, 3, 9]:

$$M_m(\xi) = \sum_{i=0}^m (m(\xi_i - \xi)_{+}^{m-1} / \omega'(\xi_i)), \quad (26)$$

where

$$\omega(\xi) = (\xi - \xi_0)(\xi - \xi_1) \dots (\xi - \xi_m) \quad (27)$$

and

$$g_+(\xi) = \begin{cases} g(\xi), & \text{if } g(\xi) \geq 0, \\ 0, & \text{if } g(\xi) < 0. \end{cases} \quad (28)$$

If we: (1) assume that the degree of the B-spline is odd, i.e., $r = m-1 = 2k-1$, k a positive integer; (2) center the spline about the origin as required by Theorems 2 and 3; and (3) let the knots of the spline occur at integer values; then we may obtain the B-spline representation for K_0 (see [9]).

$$K_0(\xi) = M_{r+1}(\xi) = \frac{1}{r!} \sum_{\ell=-k}^k (-1)^{\ell+k} \binom{2k}{\ell+k} (\xi - \xi)_{+}^r, \quad (29)$$

where, as indicated above, $m = r+1 = 2k$.

Substituting (29) in (21), (23) and then (19) and (20), we obtain the formula for optimal $h^*(N)$ and the optimal kernel-dependent rate of convergence $\Lambda_2(K_0)$:

$$h^*(N) = \frac{1}{N} \frac{1}{n+4} \left[\frac{144n(M_m(0))^n}{m^2 \left\| \sum_{i=1}^n \frac{\partial^2 f_X}{\partial x_i^2} \right\|^2} \right]^{\frac{1}{n+4}} \quad (30)$$

and

$$\Lambda_2(K_0) = \left(\frac{m}{12} \right)^{\frac{2n}{n+4}} \gamma_m^{\frac{4n}{n+4}}, \quad (31)$$

where

$$\gamma_m = \frac{1}{(2r+1)!} \sum_{j=1}^{r+1} (-1)^{j+r+1} \frac{2(r+1)}{(j+r+1)} j^{2r+1}. \quad (32)$$

Numerical values for $\Lambda_2(K_0)$ given by (31), for $r = 1, 3$, and 5 , are listed in Table I.

TABLE I

r	$\Lambda_2(K_0)$
1	.353075
3	.357836
5	.359683

Let $\eta(\mu, \Sigma; x)$ denote the value at x of the normal density with mean μ and covariance matrix Σ . For $K_0(x) = \eta(0, \sigma^2; x)$ we have

$$\|\eta^{(2)}(0, \sigma^2; \cdot)\|_2^2 = (3/8\pi^{-5}\sigma^{-5}). \quad (33)$$

Using (33), with $\sigma = 1$, in (30), we get the formulas for $h^*(N)$, for any dimension n and $r = 1, 3$, presented in Table II.

TABLE II

r	$h^*(N)$
1	$\left[\frac{36n(.66666)^n}{.2115} \right]^{\frac{1}{n+4}} N^{-\frac{1}{n+4}}$
3	$\left[\frac{9n(.49365)^n}{.2115} \right]^{\frac{1}{n+4}} N^{-\frac{1}{n+4}}$

5. Computer Simulation Results

In this section we present some of the computer simulation results performed on the Rice University IBM 370/155 digital computer for the purpose of testing how well the B-spline potential function algorithm performs in the construction of likelihood functions.

Given a set of samples $T(N) = \{y_1, \dots, y_n\}$, where each y_i is an independent realization of a random variable X with density f_X , let the sample mean $\tilde{\mu}$ and sample covariance matrix $\tilde{\Sigma}$ be defined in the usual way, i.e.

$$\tilde{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i \quad (34)$$

and

$$\tilde{\Sigma} = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{\mu})(y_i - \tilde{\mu})^T, \quad (35)$$

where the superscript T denotes the transpose. Then $\eta(\tilde{\mu}, \tilde{\Sigma}; x)$ will be called the "sample normal density".

The simulation results mentioned above are shown in Figs. 1 through 6.

Fig. 1 is a graphic display of the bimodal density

$$f_X(x) = p\eta(\mu_1, \Sigma; x) + (1-p)\eta(\mu_2, \Sigma; x) \quad (36)$$

with the mixing parameter $p = .5$, and $\mu_1 = \text{col}(-2, 0)$, $\mu_2 = \text{col}(6, 0)$ and

$$\Sigma = \begin{pmatrix} 5.75 & 4.34 \\ 4.34 & 6.64 \end{pmatrix}. \quad (37)$$

Figs. 2 and 3 are displays of $\hat{f}_X(x/T(N))$ obtained by the B-spline potential function algorithm corresponding respectively to $N=50$ and $N = 300$ samples from the above density. Fig. 4 shows the sample normal density approximation of the same density on the basis of 50 samples.

To show the effect of the increase in dimensionality on the performance of our algorithm, we present in Figs. 5 and 6 the cross-sections through the x_1 -axis of the density estimates, obtained by the B-spline potential function algorithm, of the bimodal density (36) on four- and six-dimensional spaces under the conditions given in those figures.

In all this work we used cubic B-splines with the mesh parameter $h^*(N)$ equal to the second entry in Table II.

From the above few results we conclude that the B-spline potential function algorithm appears to fare well in the construction of likelihood functions from only a modest number of samples and with densities that are not necessarily unimodal and on spaces that are not necessarily of too low a dimension.

6. Application to Remote Sensing

To test the effectiveness of the B-spline potential function algorithm for classification, discriminant functions were obtained from the likelihood functions generated by the algorithm, for Bayesian classification of agricultural crops. The algorithm was based on cubic B-splines chosen as in section 4 of this paper, and was implemented in the LARSYSA VERSION 2. This was developed at the Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Indiana. We also performed classification using sample normal densities as likelihood functions. (See the beginning of section 5 for the definition of "sample normal density".)

The data used in our experiments pertained to the Purdue LARS flight line C1, which has been widely employed for testing algorithms on remote sensing. This data consists of the output of a twelve channel spectrometer which analyzes the reflected radiance from the object being

sensed. Let a given crop field belonging, say, to the pattern class H^j , be discretized (partitioned) into N_j points, called "resolution elements". The spectrometer maps each resolution element into a 12-tuple of real numbers, say $y_k^{(j)} = \text{col}(y_{k1}^{(j)}, \dots, y_{k12}^{(j)})$, and the whole field provides N_j such 12-dimensional vector samples $y_1^{(j)}, \dots, y_{N_j}^{(j)}$, which can be used to train a classifier in the acquisition of the likelihood function corresponding to H^j .

Our first example is designed to show the poor results obtained when normality is assumed on bimodal data. In this example, we used data from only one channel, namely Channel 1 (.40 μ to .44 μ), to classify data corresponding to the two bimodal pattern classes: H^1 : RED CLOVER HAY and CORN1; and H^2 : BARE SOIL1 and ALFALFA1. Figs. 7 and 8 show the histograms of the classes. The percentages of the number of correct classifications, for a typical set of observations corresponding to H^1 and H^2 , are indicated in Table III, both for the algorithm presented here and for the sample normal classification algorithm.

TABLE III

Potential Function Algorithm	Sample Normal Algorithm
H^1	86%
H^2	98%

It is clear that the much superior performance of the potential function algorithm in relation to the sample normal algorithm may be attributed to the bi-modality of the data.

Our second example is for the purpose of testing the effectiveness of the potential function algorithm for normal data. In this example, we used 3 channels, namely Channels 1 (.40 μ - .44 μ), 10 (.66 μ - .72 μ), and 12 (.60 μ - 1.00 μ), to classify H^1 : SOYBEANS, H^2 : CORN, H^3 : OATS, and H^4 : WHEAT. The percentages of correct classifications are displayed in Table IV, for H^1 and H^2 .

TABLE IV

Potential Function Algorithm	Sample Normal Algorithm
H^1	97%
H^2	94%

Even though the efficiency of classification by the potential function algorithm is lower than by the sample normal, we note that the ability of the potential function algorithm to classify effectively data that is normal is comparable with the sample normal in quality of classification.

Acknowledgement

We are much indebted to Dr. D. Van Rooy for implementing the classification algorithm into the LARNSYSA system and to Mr. Ken Baker of the Lyndon B. Johnson Space Center for his advice on the remote sensing problem investigated. It

is also a pleasure to acknowledge a helpful comment from Professor Polking on the proof of Theorem 1 and the encouragement received in carrying out this research from Dr. M. S. Lynn, Director of the Rice University Institute for Computer Services and Applications.

Appendix

Proof of Theorem 1

For notational convenience let

$$K_h(x) = h^{-n} K(h^{-1}x). \quad (A-1)$$

Then, clearly* (see for example [5, p. 172]),

$$\begin{aligned} & E\{\hat{f}_X(x/T(N)) - f_X(x)\}^2 \\ & = \text{Var}(\hat{f}_X(x/T(N))) + B^2(\hat{f}_X(x/T(N))), \end{aligned} \quad (A-2)$$

where

$$\begin{aligned} \text{Var}(\hat{f}_X(x/T(N))) & = E\{\hat{f}_X^2(x/T(N))\} \\ & - E^2\{\hat{f}_X(x/T(N))\} \end{aligned} \quad (A-3)$$

and

$$B(\hat{f}_X(x/T(N))) = E\{\hat{f}_X(x/T(N))\} - f_X(x). \quad (A-4)$$

Hence it follows that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \|E\{\hat{f}_X(x/T(N)) - f_X(x)\}^2\|_1 \\ & = \lim_{N \rightarrow \infty} \|\text{Var}(\hat{f}_X(x/T(N)))\|_1 \\ & + \lim_{N \rightarrow \infty} \|B(\hat{f}_X(x/T(N)))\|_2^2. \end{aligned} \quad (A-5)$$

The proof will consist of showing that each term on the right side of (A-5) tends to zero.

Since the random variables Y_i , $i = 1, \dots, N$ are each independently distributed as X , we have, in accordance with (3), that

$$E\{\hat{f}_X(x/T(N))\} = E\{K_h(x - X)\}, \quad (A-6)$$

and

$$\text{Var}\{\hat{f}_X^2(x/T(N))\} = N^{-1} \text{Var}\{K_h(x - X)\}. \quad (A-7)$$

Now

$$\begin{aligned} K_h(x) * f_X(x) & = \int_{\mathbb{R}^n} K_h(x - z) f_X(z) dz \\ & = E\{K_h(x - X)\} \\ & = E\{\hat{f}_X(x/T(N))\}, \end{aligned} \quad (A-8)$$

where, in going from the third to the last member, we have used (A-6).

Similarly,

*In this Appendix, we use capital letters for symbols denoting random variables and corresponding small letters for realizations of these random variables. In (A-2), $\hat{f}_X(x/T(N))$ is to be regarded as a function of the random variables Y_1, \dots, Y_N the realizations of which are the training samples y_1, \dots, y_N .

$$K_h^2(x) * f_X(x) = E[K_h^2(x-X)] . \quad (A-9)$$

Hence, from (A-7), using (A-8) and (A-9) we get

$$\begin{aligned} & \|\text{Var}(\hat{f}_X(x/T(N)))\|_1 = \int_{\mathbb{R}^n} \text{Var}(f_X(x/T(N))) dx \\ &= \int_{\mathbb{R}^n} [E\{K_h^2(x-X)\} - E^2\{K_h(x-X)\}] dx \\ &= \|E\{K_h^2(x-X)\}\|_1 - \|E\{K_h(x-X)\}\|_2^2 \\ &= (Nh^n)^{-1} \|(K^2(x))_h * f_X(x)\|_1 \\ &\quad - N^{-1} \|K_h(x) * f_X(x)\|_2^2 , \end{aligned} \quad (A-10)$$

where

$$(K^2(x))_h = h^{-n} K^2(h^{-1}x) . \quad (A-11)$$

By Young's inequality (see [8], p. 148), we have

$$\|(K^2(x))_h * f_X(x)\|_2^2 \leq \|f_X(x)\|_2^2 \quad (A-12)$$

since

$$\|K_h(x)\|_1^2 , \quad (A-13)$$

and again by Young's inequality

$$\|(K^2(x))_h * f_X(x)\|_1 \leq \|(K^2(x))_h\|_1 \quad (A-14)$$

since

$$\|f_X(x)\|_1 = 1 . \quad (A-15)$$

By a change of variables we obtain

$$\|(K^2(x))_h\|_1 = \|K(x)\|_2^2 , \quad (A-16)$$

and so (A-14) becomes

$$\|(K^2(x))_h * f_X(x)\|_1 \leq \|K(x)\|_2^2 . \quad (A-17)$$

Using the triangle inequality on the right side of (A-10), and then substituting into it (A-17) and (A-12), we obtain

$$\begin{aligned} \|\text{Var}(\hat{f}_X(x/T(N)))\|_1 &\leq (Nh^n)^{-1} \|K(x)\|_2^2 \\ &\quad + N^{-1} \|f_X(x)\|_2^2 . \end{aligned} \quad (A-18)$$

Finally, resorting to the hypotheses II, III-1, and IV-2 of the theorem, we have

$$\lim_{N \rightarrow \infty} \|\text{Var}(\hat{f}_X(x/T(N)))\|_1 = 0 . \quad (A-19)$$

Now consider the bias term (A-4) and use (A-8) to write it in the form

$$B(f_X(x|T(N))) = K_h(x) * f_X(x) - f_X(x) . \quad (A-20)$$

Then by Theorem 2, Part (c) in Stein [11, p.62] we get

$$\lim_{N \rightarrow \infty} \|B(\hat{f}_X(x/T(N)))\|_2 = 0 . \quad (A-21)$$

Equations (A-5), (A-19), and (A-21) show that $\hat{f}_X(x/T(N))$ is a consistent estimator of $f_X(x)$ in the IMS sense.

BIBLIOGRAPHY

- [1] Cacoullos, T., Estimation of a Multivariate Density, Annals of the Institute of Statistical Mathematics (Tokyo), Volume 18, pp. 179-189, 1966.
- [2] Curry, H.B., and Schoenberg, I.J., On Polya Frequency Functions IV. The Fundamental Spline Functions and their Limits, Journal d'Analyse Mathematique (Jerusalem), Volume 17, pp. 71-107, 1966.
- [3] de Boor, C., Package for Calculating with B-splines, Mathematics Research Center (Madison, Wisconsin), Technical Summary Report #1333 (April), 1973.
- [4] Fu, K.S., Sequential Methods in Pattern Recognition and Machine Learning, Academic Press (New York) 1968.
- [5] Mood, A.M., and Graybill, F.A., Introduction to the Theory of Statistics, McGraw-Hill (New York) 1963.
- [6] Murthy, V.K., Non-Parametric Estimation of Multivariate Densities with Applications, Multivariate Analysis II, International Symposium on Multivariate Analysis at Wright State University (Dayton, Ohio), Academic Press (New York), 1966.
- [7] Parzen, E., On Estimation of Probability Density Function and Mode, The Annals of Mathematical Statistics, (Baltimore, Maryland) Volume 33, # 3 (September), pp. 1065-1076, 1962.
- [8] Rudin, W., Real and Complex Analysis, McGraw-Hill Book Company (New York), 1966.
- [9] Schoenberg, I.J., Cardinal Interpolation and Spline Functions, Journal of Approximation Theory, (New York), Volume 2, pp. 167-206, 1969.
- [10] Specht, D.F., Generation of Polynomial Discriminant Functions for Pattern Recognition, Institute of Electrical and Electronic Engineers Transactions on Electronic Computers (New York), Volume EC-16, # 3, (June) 1967.
- [11] Stein, E.M., Singular Integrals and Differentiability Properties of Functions, Princeton University Press (Princeton, N.J.), 1970.

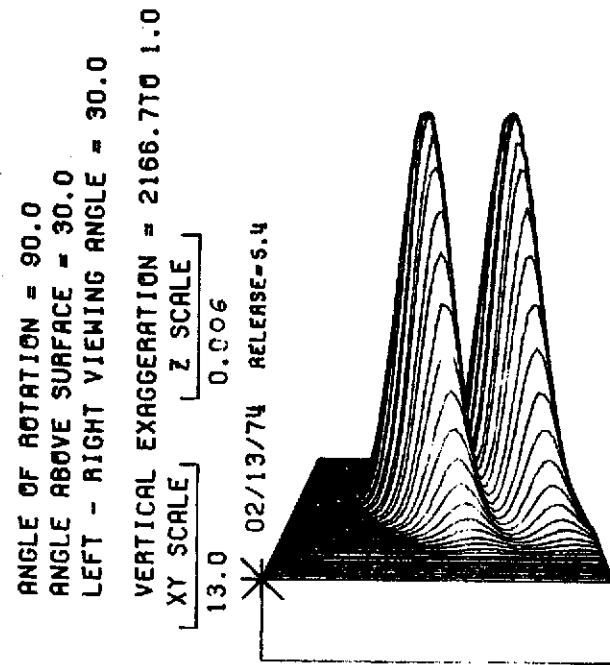


FIG.1 PERSPECTIVE VIEW OF
 A BIMODAL DENSITY

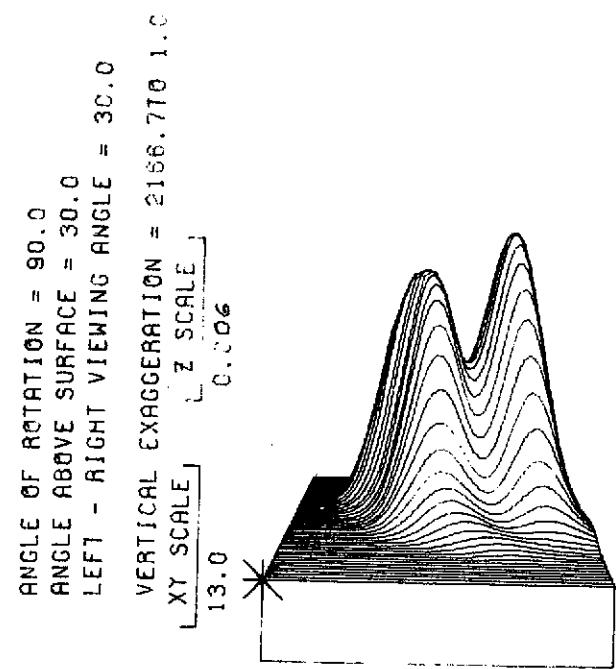


FIG.3 DENSITY KERNEL ESTIMATOR
 SAMPLE SIZE = 300. CUBIC B-SPLINE BASE

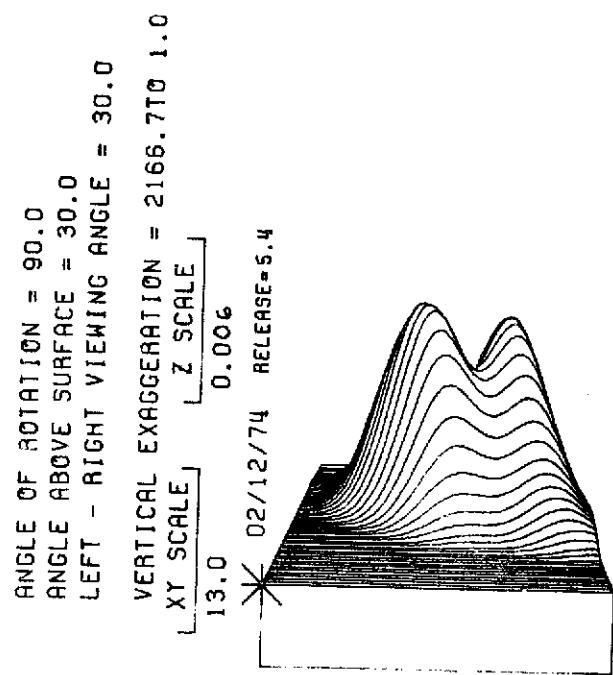


FIG.2 PERSPECTIVE VIEW OF
 DENSITY KERNEL ESTIMATOR,
 SAMPLE SIZE=50, CUBIC B-SPLINE

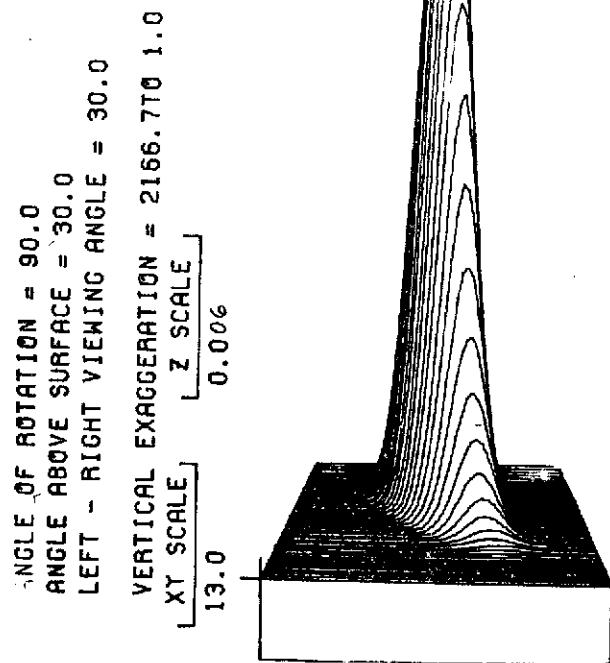


FIG.4 PERSPECTIVE VIEW OF SAMPLE NORMAL
 WITH SAMPLE SIZE = 50

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

APPROXIMATION *** NORMAL DENSITY 0.00 CROSS SECTION WITH ALL COORDINATES BUT X1 = 1
 SET TU 0.0 DIMENSION IS 4 SAMPLE SIZE = 100 ALPHA = -3+1250 M = 1.131656
 ABSCISSA UNDERRATE MIN
 ABSISSA 0.0
 0.0
 0.1
 0.2
 0.3
 0.4
 0.5
 0.6
 0.7
 0.8
 0.9
 1.0
 1.1
 1.2
 1.3
 1.4
 1.5
 1.6
 1.7
 1.8
 1.9
 2.0
 2.1
 2.2
 2.3
 2.4
 2.5
 2.6
 2.7
 2.8
 2.9
 3.0
 3.1
 3.2
 3.3
 3.4
 3.5
 3.6
 3.7
 3.8
 3.9
 4.0
 4.1
 4.2
 4.3
 4.4
 4.5
 4.6
 4.7
 4.8
 4.9
 5.0
 5.1
 5.2
 5.3
 5.4
 5.5
 5.6
 5.7
 5.8
 5.9
 6.0
 6.1
 6.2
 6.3
 6.4
 6.5
 6.6
 6.7
 6.8
 6.9
 7.0
 7.1
 7.2
 7.3
 7.4
 7.5
 7.6
 7.7
 7.8
 7.9
 8.0
 8.1
 8.2
 8.3
 8.4
 8.5
 8.6
 8.7
 8.8
 8.9
 9.0
 9.1
 9.2
 9.3
 9.4
 9.5
 9.6
 9.7
 9.8
 9.9
 10.0
 10.1
 10.2
 10.3
 10.4
 10.5
 10.6
 10.7
 10.8
 10.9
 11.0
 11.1
 11.2
 11.3
 11.4
 11.5
 11.6
 11.7
 11.8
 11.9
 12.0
 12.1
 12.2
 12.3
 12.4
 12.5
 12.6
 12.7
 12.8
 12.9
 13.0
 13.1
 13.2
 13.3
 13.4
 13.5
 13.6
 13.7
 13.8
 13.9
 14.0
 14.1
 14.2
 14.3
 14.4
 14.5
 14.6
 14.7
 14.8
 14.9
 15.0
 15.1
 15.2
 15.3
 15.4
 15.5
 15.6
 15.7
 15.8
 15.9
 16.0
 16.1
 16.2
 16.3
 16.4
 16.5
 16.6
 16.7
 16.8
 16.9
 17.0
 17.1
 17.2
 17.3
 17.4
 17.5
 17.6
 17.7
 17.8
 17.9
 18.0
 18.1
 18.2
 18.3
 18.4
 18.5
 18.6
 18.7
 18.8
 18.9
 19.0
 19.1
 19.2
 19.3
 19.4
 19.5
 19.6
 19.7
 19.8
 19.9
 20.0
 20.1
 20.2
 20.3
 20.4
 20.5
 20.6
 20.7
 20.8
 20.9
 21.0
 21.1
 21.2
 21.3
 21.4
 21.5
 21.6
 21.7
 21.8
 21.9
 22.0
 22.1
 22.2
 22.3
 22.4
 22.5
 22.6
 22.7
 22.8
 22.9
 23.0
 23.1
 23.2
 23.3
 23.4
 23.5
 23.6
 23.7
 23.8
 23.9
 24.0
 24.1
 24.2
 24.3
 24.4
 24.5
 24.6
 24.7
 24.8
 24.9
 25.0
 25.1
 25.2
 25.3
 25.4
 25.5
 25.6
 25.7
 25.8
 25.9
 26.0
 26.1
 26.2
 26.3
 26.4
 26.5
 26.6
 26.7
 26.8
 26.9
 27.0
 27.1
 27.2
 27.3
 27.4
 27.5
 27.6
 27.7
 27.8
 27.9
 27.10
 27.11
 27.12
 27.13
 27.14
 27.15
 27.16
 27.17
 27.18
 27.19
 27.20
 27.21
 27.22
 27.23
 27.24
 27.25
 27.26
 27.27
 27.28
 27.29
 27.30
 27.31
 27.32
 27.33
 27.34
 27.35
 27.36
 27.37
 27.38
 27.39
 27.40
 27.41
 27.42
 27.43
 27.44
 27.45
 27.46
 27.47
 27.48
 27.49
 27.50
 27.51
 27.52
 27.53
 27.54
 27.55
 27.56
 27.57
 27.58
 27.59
 27.60
 27.61
 27.62
 27.63
 27.64
 27.65
 27.66
 27.67
 27.68
 27.69
 27.70
 27.71
 27.72
 27.73
 27.74
 27.75
 27.76
 27.77
 27.78
 27.79
 27.80
 27.81
 27.82
 27.83
 27.84
 27.85
 27.86
 27.87
 27.88
 27.89
 27.90
 27.91
 27.92
 27.93
 27.94
 27.95
 27.96
 27.97
 27.98
 27.99
 27.100
 27.101
 27.102
 27.103
 27.104
 27.105
 27.106
 27.107
 27.108
 27.109
 27.110
 27.111
 27.112
 27.113
 27.114
 27.115
 27.116
 27.117
 27.118
 27.119
 27.120
 27.121
 27.122
 27.123
 27.124
 27.125
 27.126
 27.127
 27.128
 27.129
 27.130
 27.131
 27.132
 27.133
 27.134
 27.135
 27.136
 27.137
 27.138
 27.139
 27.140
 27.141
 27.142
 27.143
 27.144
 27.145
 27.146
 27.147
 27.148
 27.149
 27.150
 27.151
 27.152
 27.153
 27.154
 27.155
 27.156
 27.157
 27.158
 27.159
 27.160
 27.161
 27.162
 27.163
 27.164
 27.165
 27.166
 27.167
 27.168
 27.169
 27.170
 27.171
 27.172
 27.173
 27.174
 27.175
 27.176
 27.177
 27.178
 27.179
 27.180
 27.181
 27.182
 27.183
 27.184
 27.185
 27.186
 27.187
 27.188
 27.189
 27.190
 27.191
 27.192
 27.193
 27.194
 27.195
 27.196
 27.197
 27.198
 27.199
 27.200
 27.201
 27.202
 27.203
 27.204
 27.205
 27.206
 27.207
 27.208
 27.209
 27.210
 27.211
 27.212
 27.213
 27.214
 27.215
 27.216
 27.217
 27.218
 27.219
 27.220
 27.221
 27.222
 27.223
 27.224
 27.225
 27.226
 27.227
 27.228
 27.229
 27.230
 27.231
 27.232
 27.233
 27.234
 27.235
 27.236
 27.237
 27.238
 27.239
 27.240
 27.241
 27.242
 27.243
 27.244
 27.245
 27.246
 27.247
 27.248
 27.249
 27.250
 27.251
 27.252
 27.253
 27.254
 27.255
 27.256
 27.257
 27.258
 27.259
 27.260
 27.261
 27.262
 27.263
 27.264
 27.265
 27.266
 27.267
 27.268
 27.269
 27.270
 27.271
 27.272
 27.273
 27.274
 27.275
 27.276
 27.277
 27.278
 27.279
 27.280
 27.281
 27.282
 27.283
 27.284
 27.285
 27.286
 27.287
 27.288
 27.289
 27.290
 27.291
 27.292
 27.293
 27.294
 27.295
 27.296
 27.297
 27.298
 27.299
 27.300
 27.301
 27.302
 27.303
 27.304
 27.305
 27.306
 27.307
 27.308
 27.309
 27.310
 27.311
 27.312
 27.313
 27.314
 27.315
 27.316
 27.317
 27.318
 27.319
 27.320
 27.321
 27.322
 27.323
 27.324
 27.325
 27.326
 27.327
 27.328
 27.329
 27.330
 27.331
 27.332
 27.333
 27.334
 27.335
 27.336
 27.337
 27.338
 27.339
 27.340
 27.341
 27.342
 27.343
 27.344
 27.345
 27.346
 27.347
 27.348
 27.349
 27.350
 27.351
 27.352
 27.353
 27.354
 27.355
 27.356
 27.357
 27.358
 27.359
 27.360
 27.361
 27.362
 27.363
 27.364
 27.365
 27.366
 27.367
 27.368
 27.369
 27.370
 27.371
 27.372
 27.373
 27.374
 27.375
 27.376
 27.377
 27.378
 27.379
 27.380
 27.381
 27.382
 27.383
 27.384
 27.385
 27.386
 27.387
 27.388
 27.389
 27.390
 27.391
 27.392
 27.393
 27.394
 27.395
 27.396
 27.397
 27.398
 27.399
 27.400
 27.401
 27.402
 27.403
 27.404
 27.405
 27.406
 27.407
 27.408
 27.409
 27.410
 27.411
 27.412
 27.413
 27.414
 27.415
 27.416
 27.417
 27.418
 27.419
 27.420
 27.421
 27.422
 27.423
 27.424
 27.425
 27.426
 27.427
 27.428
 27.429
 27.430
 27.431
 27.432
 27.433
 27.434
 27.435
 27.436
 27.437
 27.438
 27.439
 27.440
 27.441
 27.442
 27.443
 27.444
 27.445
 27.446
 27.447
 27.448
 27.449
 27.450
 27.451
 27.452
 27.453
 27.454
 27.455
 27.456
 27.457
 27.458
 27.459
 27.460
 27.461
 27.462
 27.463
 27.464
 27.465
 27.466
 27.467
 27.468
 27.469
 27.470
 27.471
 27.472
 27.473
 27.474
 27.475
 27.476
 27.477
 27.478
 27.479
 27.480
 27.481
 27.482
 27.483
 27.484
 27.485
 27.486
 27.487
 27.488
 27.489
 27.490
 27.491
 27.492
 27.493
 27.494
 27.495
 27.496
 27.497
 27.498
 27.499
 27.500
 27.501
 27.502
 27.503
 27.504
 27.505
 27.506
 27.507
 27.508
 27.509
 27.510
 27.511
 27.512
 27.513
 27.514
 27.515
 27.516
 27.517
 27.518
 27.519
 27.520
 27.521
 27.522
 27.523
 27.524
 27.525
 27.526
 27.527
 27.528
 27.529
 27.530
 27.531
 27.532
 27.533
 27.534
 27.535
 27.536
 27.537
 27.538
 27.539
 27.540
 27.541
 27.542
 27.543
 27.544
 27.545
 27.546
 27.547
 27.548
 27.549
 27.550
 27.551
 27.552
 27.553
 27.554
 27.555
 27.556
 27.557
 27.558
 27.559
 27.560
 27.561
 27.562
 27.563
 27.564
 27.565
 27.566
 27.567
 27.568
 27.569
 27.570
 27.571
 27.572
 27.573
 27.574
 27.575
 27.576
 27.577
 27.578
 27.579
 27.580
 27.581
 27.582
 27.583
 27.584
 27.585
 27.586
 27.587
 27.588
 27.589
 27.590
 27.591
 27.592
 27.593
 27.594
 27.595
 27.596
 27.597
 27.598
 27.599
 27.600
 27.601
 27.602
 27.603
 27.604
 27.605
 27.606
 27.607
 27.608
 27.609
 27.610
 27.611
 27.612
 27.613
 27.614
 27.615
 27.616
 27.617
 27.618
 27.619
 27.620
 27.621
 27.622
 27.623
 27.624
 27.625
 27.626
 27.627
 27.628
 27.629
 27.630
 27.631
 27.632
 27.633
 27.634
 27.635
 27.636
 27.637
 27.638
 27.639
 27.640
 27.641
 27.642
 27.643
 27.644
 27.645
 27.646
 27.647
 27.648
 27.649
 27.650
 27.651
 27.652
 27.653
 27.654
 27.655
 27.656
 27.657
 27.658
 27.659
 27.660
 27.661
 27.662
 27.663
 27.664
 27.665
 27.666
 27.667
 27.668
 27.669
 27.670
 27.671
 27.672
 27.673
 27.674
 27.675
 27.676
 27.677
 27.678
 27.679
 27.680
 27.681
 27.682
 27.683
 27.684
 27.685
 27.686
 27.687
 27.688
 27.689
 27.690
 27.691
 27.692
 27.693
 27.694
 27.695
 27.696
 27.697
 27.698
 27.699
 27.700
 27.701
 27.702
 27.703
 27.704
 27.705
 27.706
 27.707
 27.708
 27.709
 27.710
 27.711
 27.712
 27.713
 27.714
 27.715
 27.716
 27.717
 27.718
 27.719
 27.720
 27.721
 27.722
 27.723
 27.724
 27.725
 27.726
 27.727
 27.728
 27.729
 27.730
 27.731
 27.732
 27.733
 27.734
 27.735
 27.736
 27.737
 27.738
 27.739
 27.740
 27.741
 27.742
 27.743
 27.744
 27.745
 27.746
 27.747
 27.748
 27.749
 27.750
 27.751
 27.752
 27.753
 27.754
 27.755
 27.756
 27.757
 27.758
 27.759
 27.760
 27.761
 27.762
 27.763
 27.764
 27.765
 27.766
 27.767
 27.768
 27.769
 27.770
 27.771
 27.772
 27.773
 27.774
 27.775
 27.776
 27.777
 27.778
 27.779
 27.780
 27.781
 27.782
 27.783
 27.784
 27.785
 27.786
 27.787
 27.788
 27.789
 27.790
 27.791
 27.792
 27.793
 27.794
 27.795
 27.796
 27.797
 27.798
 27.799
 27.800
 27.801
 27.802
 27.803
 27.804
 27.805
 27.806
 27.807
 27.808
 27.809
 27.810
 27.811
 27.812
 27.813
 27.814
 27.815
 27.816
 27.817
 27.818
 27.819
 27.820
 27.821
 27.822
 27.823
 27.824
 27.825
 27.826
 27.827
 27.828
 27.829
 27.830
 27.831
 27.832
 27.833
 27.834
 27.835
 27.836
 27.837
 27.838
 27.839
 27.840
 27.841
 27.842
 27.843
 27.844
 27.845
 27.846
 27.847
 27.848
 27.849
 27.850
 27.851
 27.852
 27.853
 27.854
 27.855
 27.856
 27.857
 27.858
 27.859
 27.860
 27.861
 27.862
 27.863
 27.864
 27.865
 27.866
 27.867
 27.868
 27.869
 27.870
 27.871
 27.872
 27.873
 27.874
 27.875
 27.876
 27.877
 27.878
 27.879
 27.880
 27.881
 27.882
 27.883
 27.884
 27.885
 27.886
 27.887
 27.888
 27.889
 27.890
 27.891
 27.892
 27.893
 27.894
 27.895
 27.896
 27.897
 27.898
 27.899
 27.900
 27.901
 27.902
 27.903
 27.904
 27.905
 27.906
 27.907
 27.908
 27.909
 27.910
 27.911
 27.912
 27.913
 27.914
 27.915
 27.916
 27.917
 27.918
 27.919
 27.920
 27.921
 27.922
 27.923
 27.924
 27.925
 27.926
 27.9